REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release, distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report #88-16C	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Purdue University 6b. OFFICE SYMBOL *(If applicable)*	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS *(City, State, and ZIP Code)* Department of Statistics West Lafayette, IN 47907	7b. ADDRESS *(City, State, and ZIP Code)*
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL *(If applicable)*	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-88-K-0170, DMS-8606964, DMS-8702620
8c. ADDRESS *(City, State, and ZIP Code)* Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. | PROJECT NO. | TASK NO. | WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*
On Empirical Bayes Selection Rules for Negative Binomial Populations (Unclassified)

12. PERSONAL AUTHOR(S)
Shanti S. Gupta and TaChen Liang

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT *(Year, Month, Day)*	15. PAGE COUNT
Technical	FROM ______ TO ______	May 1988	27

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
FIELD	GROUP	SUB-GROUP	Asymptotically optimal, Bayes rules, empirical Bayes rules, monotone estimators, monotone selection rules, prior distribution, rate of convergence, antitonic regression.

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This paper deals with the problem of selecting good negative binomial populations as compared with a standard or a control. The main results are based on the use of the empirical Bayes approach. First we derive the monotone empirical Bayes estimators of the censored parameters. Based on these estimators, we construct monotone empirical Bayes selection rules. Asymptotic optimality properties of the monotone empirical Bayes estimators and the monotone empirical Bayes selection rules are investigated. The respective convergence rates for the estimation problem and for the selection problem are studied, under some conditions.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Shanti S. Gupta	22b. TELEPHONE *(Include Area Code)* 317-494-6031 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

ON EMPIRICAL BAYES SELECTION RULES FOR NEGATIVE BINOMIAL POPULATIONS*

by

Shanti S. Gupta
Purdue University

TaChen Liang
Wayne State University

Technical Report #88-16C

Department of Statistics
Purdue University

May 1988

* This research was supported in part by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8606964, DMS-8702620 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

o

ON EMPIRICAL BAYES SELECTION RULES FOR NEGATIVE BINOMIAL POPULATIONS*

by

Shanti S. Gupta
Purdue University

TaChen Liang
Wayne State University

Abstract

This paper deals with the problem of selecting good negative binomial populations as compared with a standard or a control. The main results are based on the use of the empirical Bayes approach. First we derive the monotone empirical Bayes estimators of the concerned parameters. Based on these estimators, we construct monotone empirical Bayes selection rules. Asymptotic optimality properties of the monotone empirical Bayes estimators and the monotone empirical Bayes selection rules are investigated. The respective convergence rates for the estimation problem and for the selection problem are studied, under some conditions.

AMS 1980 Subject Classification: Primary 62F07; Secondary 62C12.

Key words and phrases: Asymptotically optimal, Bayes rules, empirical Bayes rules, monotone estimators, monotone selection rules, prior distribution, rate of convergence, antitonic regression.

* This research was supported in part by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS–8606964, DMS–8702620 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. Introduction

The empirical Bayes approach in statistical decision theory is appropriate when one is confronted repeatedly and independently with the same decision problem. In such instances, it is reasonable to formulate the component problem in the sequence as Bayes decision problems with respect to an unknown prior distribution on the parameter space, and then use the accumulated observations to improve the decision rule at each stage. This approach is due to Robbins (1956, 1964, 1983). During the last twenty-five years, empirical Bayes methods have been studied extensively. Many such empirical Bayes rules have been shown to be asymptotically optimal in the sense that the risk for the nth decision problem converges to the minimum Bayes risk which would have been obtained if the prior distribution was known and the Bayes rule with respect to this prior distribution was used.

Empirical Bayes rules have been derived for subset selection goals by Deely (1965). Recently, Gupta and Hsiao (1983) and Gupta and Leu (1983) have studied empirical Bayes rules for selecting good populations with respect to a standard or a control with the underlying distributions being uniformly distributed. Gupta and Liang (1986a, 1987) studied empirical Bayes rules for selecting binomial populations better than a standard or a control and for selecting the best among several binomial populations. In the above-mentioned papers, the authors have assumed that the form of the prior distribution is completely unknown. Hence, those approaches are referred to as nonparametric empirical Bayes. Gupta and Liang (1986b, 1988) have also studied some other empirical Bayes selection rules, in which they assumed that the form of the prior distributions is known but the distributions depend on certain unknown hyperparameters. Such approach is therefore referred to as parametric empirical Bayes.

In this paper, we are concerned with the problem of selecting good negative binomial populations with respect to a standard or a control through the nonparametric empirical Bayes approach. The framework of the empirical Bayes selection problem is formulated in Section 2. Monotone empirical Bayes selection rules are proposed in Section 3. The monotone empirical Bayes selection rules are derived based on certain monotone empirical Bayes estimators of the posterior means of the concerned parameters which are also derived in Section 3. Asymptotic optimality properties of the monotone empirical Bayes estimators and the monotone empirical Bayes selection rules are studied in Section 4 and Section 5, respectively.

2. Formulation of the Empirical Bayes Approach

Consider $k+1$ independent negative binomial populations π_i, $i = 0, 1, \ldots, k$. For each i, $i = 0, 1, \ldots, k$, let p_i denote the probability of success for each trial in π_i and let X_i denote the number of successes before attaining the rth failure in π_i. We assume that the trials for each π_i are independent, $i = 0, 1, \ldots, k$. Then, conditional on p_i, $X_i|p_i$ has a negative binomial distribution with probability function $f_i(x|p_i)$ of the form

$$f_i(x|p_i) = \binom{x+r-1}{r-1} p_i^x (1-p_i)^r, \; x = 0, 1, 2, \ldots \tag{2.1}$$

Let π_0 be the control population. For each $i = 1, \ldots, k$, population π_i is said to be good if $p_i \geq p_0$ and to be bad if $p_i < p_0$, where the control parameter p_0 is either known or unknown. Our goal is to derive empirical Bayes rules to select all good populations and exclude all bad populations.

When the control parameter p_0 is known, the empirical Bayes framework can be formulated as follows:

(1) Let $\Omega = \{\underset{\sim}{p} | \underset{\sim}{p} = (p_1, \ldots, p_k),\ p_i \in (0,1) \text{ for } i = 1, \ldots, k\}$ be the parameter space. For each $\underset{\sim}{p} \in \Omega$, define $A(\underset{\sim}{p}) = \{i | p_i \geq p_0\}$ and $B(\underset{\sim}{p}) = \{i | p_i < p_0\}$. That is, $A(\underset{\sim}{p})(B(\underset{\sim}{p}))$ is the set of indices of good (bad) populations.

(2) Let $\mathcal{A} = \{a | a \subset \{1, 2, \ldots, k\}\}$ be the action space. When action a is taken, it means that population π_i is selected as a good population if $i \epsilon a$ and excluded as a bad population if $i \notin a$.

(3) For each (fixed) parameter $\underset{\sim}{p}$ and action a, the loss function $L(\underset{\sim}{p}, a)$ is defined as:

$$L(\underset{\sim}{p}, a) = \sum_{i \in A(\underset{\sim}{p}) \setminus a} (p_i - p_0) + \sum_{i \in a \setminus A(\underset{\sim}{p})} (p_0 - p_i), \tag{2.2}$$

where the first summation is the loss due to not selecting some good populations and the second summation is the loss due to selecting some bad populations.

(4) Let $dG(\underset{\sim}{p}) = \prod_{i=1}^{k} dG_i(p_i)$ be the prior distribution on the parameter space Ω, where $G_i(\cdot)$ are unknown for all $i = 1, \ldots, k$.

(5) For each i, $i = 1, \ldots, k$, let $(X_{ij},\ P_{ij})$, $j = 1, 2, \ldots, n$, be independent random vectors associated with population π_i, where X_{ij} is observable but P_{ij} is not observable. P_{ij} has prior distribution G_i. Conditional on $P_{ij} = p_{ij}$, $X_{ij} | p_{ij}$ has a negative binomial distribution with parameters r and p_{ij}. Let the jth stage observations be denoted by $\underset{\sim}{X}_j$. That is, $\underset{\sim}{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})$. From the assumptions, $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots \underset{\sim}{X}_n$, are mutually independent and identically distributed.

(6) Let $\underset{\sim}{X}_{n+1} = \underset{\sim}{X} = (X_1, \ldots, X_k)$ denote the present observation. Conditional on $\underset{\sim}{p} = (p_1, \ldots, p_k)$, $\underset{\sim}{X}$ has a joint probability function $f(\underset{\sim}{x} | \underset{\sim}{p}) = \prod_{i=1}^{k} f_i(x_i | p_i)$, where $\underset{\sim}{x} = (x_1, \ldots, x_k)$ and where $f_i(x_i | p_i)$ is the negative binomial probability function given in (2.1) for each $i = 1, \ldots, k$.

Finally, since we are interested in Bayes rules, we can restrict our attention to the nonrandomized selection rules.

(7) Let $D = \{d | d : \mathcal{X} \to \mathcal{A}$, being measurable$\}$ be the set of nonrandomized selection rules, where $\mathcal{X} = \prod_{i=1}^{k} \{0, 1, 2, \ldots\}$ is the sample space. For each $d \in D$, let $r(G, d)$ denote the associated Bayes risk. Then $r(G) \equiv \inf_{d \in D} r(G, d)$ is the minimum Bayes risk and a rule, say d_G, is called a Bayes selection rule if $r(G, d_G) = r(G)$.

When the control parameter p_0 is unknown, the indices in the associated notations should begin at 0 instead of at 1. In the sequel, (0) will be used to show this additional fact.

We now consider decision rules $d_n(\underset{\sim}{x}; \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$ whose form depends on the present observation $\underset{\sim}{x}$ and the n past data $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$ at hand. Let $r(G, d_n)$ denote the overall Bayes risk associated with the selection rule $d_n(\underset{\sim}{x}; \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$. That is,

$$r(G, d_n) \equiv \sum_{\underset{\sim}{x} \in \mathcal{X}} E \int_{\Omega} L(\underset{\sim}{p}, d_n(\underset{\sim}{x}; \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)) f(\underset{\sim}{x} | \underset{\sim}{p}) dG(\underset{\sim}{p}) \tag{2.3}$$

where the expectation E is taken with respect to $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$. For simplicity, $d_n(\underset{\sim}{x}; \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$ will be denoted by $d_n(\underset{\sim}{x})$.

Definition 2.1. A sequence of selection rules $\{d_n(\underset{\sim}{x})\}_{n=1}^{\infty}$ is said to be asymptotically optimal (a.o.) relative to the prior distribution G if $r(G, d_n) \to r(G)$ as $n \to \infty$.

For constructing a sequence of a.o. empirical Bayes rules, we first need to find the minimum Bayes risk and the associated Bayes rule d_G. From (2.2), the Bayes risk associated with the selection rule d is:

$$\begin{aligned} r(G, d) &= \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i \in d(\underset{\sim}{x})} \int_{\Omega} (p_0 - p_i) f(\underset{\sim}{x} | \underset{\sim}{p}) dG(\underset{\sim}{p}) + C \\ &= \begin{cases} \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i \in d(\underset{\sim}{x})} [p_0 - \varphi_i(x_i)] f(\underset{\sim}{x}) + C & \text{if } p_0 \text{ is known,} \\ \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i \in d(\underset{\sim}{x})} [\varphi_0(x_0) - \varphi_i(x_i)] f(\underset{\sim}{x}) + C & \text{if } p_0 \text{ is unknown,} \end{cases} \end{aligned} \tag{2.4}$$

where $f(\underset{\sim}{x}) = \prod_{i=m}^{k} f_i(x_i)$, $m = 0(1)$ if p_0 is unknown (known),

$$f_i(x_i) = \int_0^1 f_i(x_i|p)dG_i(p) = \binom{x_i + r - 1}{r - 1} \int_0^1 p^{x_i}(1-p)^r dG_i(p)$$
$$= \beta(x_i)h_i(x_i),$$
$$\beta(x_i) = \binom{x_i + r - 1}{r - 1}, h_i(x_i) = \int_0^1 p^{x_i}(1-p)^r dG_i(p).$$
$$\varphi_i(x_i) = \frac{h_i(x_i + 1)}{h_i(x_i)} \quad (\text{ note that } 0 < \varphi_i(x_i) < 1),$$
$$C = \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=1}^{k} \int_{\Omega} (p_i - p_0) I_{(p_0,1)}(p_i) f(\underset{\sim}{x}|\underset{\sim}{p}) dG(\underset{\sim}{p})$$

and $I_A(\cdot)$ is the indicator function of the set A.

Note that in (2.4), C is a constant and does not affect the determination of the Bayes rule. Thus, the nonrandomized selection rule d_G can be obtained as follows:

$$d_G(\underset{\sim}{x}) = \{i|\Delta_{iG}(\underset{\sim}{x}) \leq 0\}, \tag{2.5}$$

where

$$\Delta_{iG}(\underset{\sim}{x}) = \begin{cases} p_0 - \varphi_i(x_i) & \text{if } p_0 \text{ is known,} \\ \varphi_0(x_0) - \varphi_i(x_i) & \text{if } p_0 \text{ is unknown.} \end{cases} \tag{2.6}$$

Now, for each $i = (0), 1, \ldots, k$, based on the past data $X_{i1}, \ldots, X_{in}$, and the present observation x_i, let $\varphi_{in}(x_i) \equiv \varphi_{in}(x_i; X_{i1}, \ldots, X_{in})$ be an estimator of $\varphi_i(x_i)$, and let

$$\Delta_{in}(\underset{\sim}{x}) = \begin{cases} p_0 - \varphi_{in}(x_i) & \text{if } p_0 \text{ is known,} \\ \varphi_{0n}(x_0) - \varphi_{in}(x_i) & \text{if } p_0 \text{ is unknown.} \end{cases} \tag{2.7}$$

We then define an empirical Bayes selection rule $d_n(\underset{\sim}{x}; \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n) \equiv d_n(\underset{\sim}{x})$ as follows:

$$d_n(\underset{\sim}{x}) = \{i|\Delta_{in}(\underset{\sim}{x}) \leq 0\}. \tag{2.8}$$

If $\varphi_{in}(x) \xrightarrow{P} \varphi_i(x)$ for all $x = 0, 1, 2, \ldots$ and $i = (0), 1, \ldots, k$, where "$\xrightarrow{P}$" means convergence in probability, then $\Delta_{in}(\underset{\sim}{x}) \xrightarrow{P} \Delta_{iG}(\underset{\sim}{x})$ for all $\underset{\sim}{x} \in \mathcal{X}$. Therefore, from Corollary 2 of Robbins (1964), it follows that $r(G, d_n) \to r(G)$ as $n \to \infty$. So, the sequence of

empirical Bayes selection rules $\{d_n(\underset{\sim}{x})\}$ defined in (2.8) is asymptotically optimal for our selection problem. Hence, in the following, we have only to find sequences of estimators $\{\varphi_{in}(x)\}$, $i = (0), 1, \ldots, k$, possessing the above mentioned convergence property.

3. The Proposed Empirical Bayes Selection Rules

Before we proceed to construct empirical Bayes estimators $\{\varphi_{in}(x)\}$, we first investigate some properties of the Bayes selecting rule d_G defined in (2.5) and (2.6).

Definition 3.1. An estimator $\varphi(\cdot)$ is called a monotone estimator if $\varphi(x)$ is an increasing function of x.

Note that for each $i = (0), 1, \ldots, k$, $\varphi_i(x) = \frac{h_i(x+1)}{h_i(x)}$. Straight computations lead to the fact that $\varphi_i(x)$ is an increasing function of $x = 0, 1, 2 \ldots$ Also, note that $\varphi_i(x)$ is the posterior mean of p_i given $X_i = x$, and it is the Bayes estimator of p_i given $X_i = x$ for squared error loss.

Let $\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$ such that $x_i \le y_i$ for $i = (0), 1, \ldots, k$.

Definition 3.2.

a) When the control parameter p_0 is known, a selection rule d is said to be monotone if $d(\underset{\sim}{x}) \subseteq d(\underset{\sim}{y})$.

b) When the control parameter p_0 is unknown, a selection rule d is said to be monotone if the following two conditions are satisfied:

(b1) If $x_0 = y_0$, then $d(\underset{\sim}{x}) \subseteq d(\underset{\sim}{y})$.

(b2) If $x_0 < y_0$ and $x_i = y_i$ for all $i = 1, \ldots, k$, then $d(\underset{\sim}{x}) \supseteq d(\underset{\sim}{y})$.

By the monotone property of the Bayes estimators $\varphi_i(x)$, $i = (0), 1, \ldots, k$, one can see that the Bayes selection rule d_G is a monotone selection rule for both cases, where the control parameter p_0 is either known or unknown.

Under the squared error loss, the problem of estimating the probabiltiy of success in a negative binomial distribution is a monotone estimation problem. By Theorem 8.7 of Berger (1985), for a monotone estimation problem, the class of monotone estimators form an essentially complete class. Also, for the present selection problem, under the linear loss given in (2.2), the problem is a monotone decision problem. Again, from Berger (1985), the class of monotone selection rules form an essentially complete class. Now, one can see that if the estimators $\varphi_{in}(x)$, $i = (0), 1, \ldots, k$, are monotone, then the empirical Bayes selection rule given in (2.7) and (2.8) is also monotone. From these considerations, it is reasonable to require that the concerned estimators $\{\varphi_{in}(x)\}$ possess the above-mentioned monotone property.

Lin (1972) studied some empirical Bayes estimation problems for a class of discrete exponential family distributions which include the negative binomial distributions as a special case. Though, the estimation procedure he proposed is asymptotically optimal, the proposed estimator does not possess the monotone property. Recently, Liang (1988) studied empirical Bayes test for the class of discrete exponential family distributions. Liang (1988) proposed a monotone empirical Bayes estimator for the corresponding Bayes estimator. Based on that monotone empirical Bayes estimator, the empirical Bayes test which he proposed is monotone. Here, we use this idea of Liang (1988) with some modification since one can obtain more information about the interesting parameter p, the probability of success, for the underlying negative binomial distribution.

The Proposed Monotone Empirical Bayes Selection Rules

Let $F_i(x)$ denote the marginal distribution function of the random variable X_i. That is,

$$F_i(x) = \sum_{y=0}^{x} f_i(y) = \sum_{y=0}^{x} \beta(y)h_i(y), \; x = 0, 1, 2, \ldots, \tag{3.1}$$

where $\beta(y) = \binom{y+r-1}{r-1}$ and $h_i(y) = \int_0^1 p^y(1-p)^r dG_i(p)$, which is decreasing in y. The form of (3.1) will be used to construct our empirical Bayes estimators.

For each $i = (0), 1, \ldots, k$, based on the past data $X_{i1}, \ldots, X_{in}$, let

$$f_{in}(x) = \frac{1}{n}\sum_{j=1}^{n} I_{\{x\}}(X_{ij}) \tag{3.2}$$

and

$$h_{in}(x) = \frac{f_{in}(x)}{\beta(x)} \tag{3.3}$$

for $x = 0, 1, 2 \ldots$.

In view of the decreasing property of the function $h_i(y)$, it is desirable that the corresponding estimator possess the same property. To meet this requirement, we let $\{h_{in}^*(x)\}_{x=0}^{\infty}$ be the antitonic regression of $\{h_{in}(x)\}_{x=0}^{\infty}$ with weight $\{\beta(x)\}_{x=0}^{\infty}$. Then, let

$$F_{in}^*(x) = \sum_{y=0}^{x} f_{in}^*(y), \tag{3.4}$$

where

$$f_{in}^*(y) = \beta(y)h_{in}^*(y), y = 0, 1, 2 \ldots. \tag{3.5}$$

Note that $h_{in}^*(x)$ is nonincreasing in x, and $h_{in}^*(x) = 0$ if $x > \max(X_{i1}, \ldots, X_{in})$ for each n and each i. By the decreasing property of the function $h_i(x)$ in x, from Barlow, et al (1972),

$$F_{in}^*(x) \geq F_{in}(x) \tag{3.6}$$

and

$$\sup_{x\geq 0}|F^*_{in}(x)-F_i(x)| \leq \sup_{x\geq 0}|F_{in}(x)-F_i(x)| \tag{3.7}$$

where $F_{in}(x)$ is the empirical distribution function based on $X_{i1},\ldots,X_{in}$. That is, $F_{in}(x) = \frac{1}{n}\sum_{j=1}^{n} I_{[0,x]}(X_{ij}) = \sum_{y=0}^{[x]} \beta(y)h_{in}(y)$, where $[x]$ denotes the largest integer not greater than x.

Now, for each $x = 0,1,2,\ldots,\max(X_{i1},\ldots,X_{in})-1$, define

$$\varphi_{in}(x) = \frac{h^*_{in}(x+1)}{h^*_{in}(x)}. \tag{3.8}$$

Note that by the nonincreasing property of the antitonic estimators $h^*_{in}(x)$, $0 \leq \varphi_{in}(x) \leq 1$ for $x = 0,1,\ldots,\max(X_{i1},\ldots,X_{in})-1$. However, $\varphi_{in}(x)$ does not possess the monotone property. A smoothed version of $\varphi_{in}(x)$, which possesses the monotone property is given as follows: Let

$$\varphi^*_{in}(x) = \begin{cases} \max\limits_{y\leq x}\varphi_{in}(y) & \text{if } y \leq \max(X_{i1},\ldots,X_{in})-1, \\ \varphi^*_{in}(\max(X_{i1},\ldots,X_{in})-1) & \text{if } y \geq \max(X_{i1},\ldots,X_{in})-1. \end{cases} \tag{3.9}$$

From (3.9), it is easy to see that $\varphi^*_{in}(x)$ possesses the monotone property. Now, for each $\underset{\sim}{x} \in \mathcal{X}$, define

$$\Delta^*_{in}(\underset{\sim}{x}) = \begin{cases} p_0 - \varphi^*_{in}(x_i) & \text{if } p_0 \text{ is known,} \\ \varphi^*_{in}(x_0) - \varphi^*_{in}(x_i) & \text{if } p_0 \text{ is unknown.} \end{cases} \tag{3.10}$$

We then popose a montone empirical Bayes selection rule, say d^*_n, as follows:

$$d^*_n(\underset{\sim}{x}) = \{i|\Delta^*_{in}(\underset{\sim}{x}) \leq 0\}. \tag{3.11}$$

Asymptotic Optimality of the Selection Rules $\{d_n^*\}$

As mentioned above, to prove the asymptotic optimality of the sequence of empirical Bayes selection rules $\{d_n^*\}$, it suffices to prove that $\varphi_{in}^*(x) \xrightarrow{P} \varphi_i(x)$ for all $x = 0, 1, 2, \ldots,$ and $i = (0), 1, \ldots, k$.

For each $x = 0, 1, 2, \ldots,$ let t be a positive number small enough so that $t + \varphi_i(x) < 1$ and $\varphi_i(x) - t > 0$. We need to prove that

$$P\{|\varphi_{in}^*(x) - \varphi_i(x)| > t\} \to 0 \text{ as } n \to \infty.$$

Now,

$$\begin{aligned} &P\{|\varphi_{in}^*(x) - \varphi_i(x)| > t\} \\ &= P\{\varphi_{in}^*(x) - \varphi_i(x) < -t\} + P\{\varphi_{in}^*(x) - \varphi_i(x) > t\}. \end{aligned} \tag{3.12}$$

By the definition of $\varphi_{in}^*(x)$, $\varphi_{in}^*(x) \geq \varphi_{in}(x)$. Thus,

$$\begin{aligned} &P\{\varphi_{in}^*(x) - \varphi_i(x) < -t\} \\ &\leq P\{\varphi_{in}(x) - \varphi_i(x) < -t\} \\ &= P\{\varphi_{in}(x) - \varphi_i(x) < -t, \text{ and } \max(X_{i1}, \ldots, X_{in}) \leq x\} \\ &\quad + P\{\varphi_{in}(x) - \varphi(x) < -t, \text{ and } \max(X_{i1}, \ldots, X_{in}) > x\}. \end{aligned} \tag{3.13}$$

We have

$$P\{\varphi_{in}(x) - \varphi_i(x) < -t \text{ and } \max(X_{i1}, \ldots, X_{in}) \leq x\} \leq [F_i(x)]^n$$

which tends to 0 as $n \to \infty$, since $0 \leq F_i(x) < 1$ for all x.

Also, note that $h_{in}^*(x) > 0$ as $x < \max(X_{i1}, \ldots, X_{in})$. Let $p(x,t) = \frac{f_i(x+1)}{f_i(x)} - \frac{t\beta(x+1)}{\beta(x)}$ and $q(x,t) = \frac{t\beta(x+1)f_i(x)}{\beta(x)}$. Note that $p(x,t) > 0$ since t is a positive number such that

$\varphi_i(x) - t > 0$. From (3.1), (3.4), (3.5), (3.7) and (3.8), we have:

$$
\begin{aligned}
&P\{\varphi_{in}(x) - \varphi_i(x) < -t \text{ and } \max(X_{i1}, \ldots, X_{in}) > x\} \\
&\le P\{h^*_{in}(x+1) - h^*_{in}(x)[\varphi_i(x) - t] < 0\} \\
&= P\{f^*_{in}(x+1) - f^*_{in}(x)p(x,t) < 0\} \qquad (3.14) \\
&= P\{F^*_{in}(x+1) - F^*_{in}(x)[1 + p(x,t)] + F^*_{in}(x-1)p(x,t) < 0\} \\
&= P\{[F^*_{in}(x+1) - F_i(x+1)] - [F^*_{in}(x) - F_i(x)][1 + p(x,t)] \\
&\qquad + [F^*_{in}(x-1) - F_i(x-1)]p(x,t) < -q(x,t)\} \\
&\le P\{F^*_{in}(x+1) - F_i(x+1) < -\frac{q(x,t)}{3}\} \\
&\qquad + P\{F^*_{in}(x) - F_i(x) > -\frac{q(x,t)}{3[1+p(x,t)]}\} \\
&\qquad + P\{F^*_{in}(x-1) - F_i(x-1) < -\frac{q(x,t)}{3p(x,t)}\} \\
&\le 3P\{\sup_{y \ge 0} |F^*_{in}(y) - F_i(y)| \ge \frac{q(x,t)}{3[1+p(x,t)]}\} \\
&\le 3P\{\sup_{y \ge 0} |F_{in}(y) - F_i(y)| \ge \frac{q(x,t)}{3[1+p(x,t)]}\}
\end{aligned}
$$

which tends to zero as n tends to infinity, by the Glivenko-Cantelli Theorem.

Similarly,

$$
\begin{aligned}
&P\{\varphi^*_{in}(x) - \varphi_i(x) > t\} \\
&= P\{\varphi^*_{in}(x) - \varphi_i(x) > t \text{ and } \max(X_{i1}, \ldots, X_{in}) \le x\} \qquad (3.15) \\
&\qquad + P\{\varphi^*_{in}(x) - \varphi_i(x) > t \text{ and } \max(X_{i1}, \ldots, X_{in}) > x\}
\end{aligned}
$$

where

$$P\{\varphi^*_{in}(x) - \varphi_i(x) > t \text{ and } \max(X_{i1}, \ldots, X_{in}) \le x\} \to 0 \text{ as } n \to \infty,$$

and by the fact that $\varphi_i(x) \geq \varphi_i(y)$ for all $y \leq x$, and the definition of $\varphi^*_{in}(x)$,

$$\begin{aligned}
&P\{\varphi^*_{in}(x) - \varphi_i(x) > t \text{ and } \max(X_{i1}, \ldots, X_{in}) > x\} \\
&= P\{\varphi_{in}(y) - \varphi_i(x) > t \text{ for some } y \leq x \text{ and } \max(X_{i1}, \ldots, X_{in}) > x\} \\
&\leq P\{\varphi_{in}(y) - \varphi_i(y) > t \text{ for some } y \leq x, f^*_{in}(y) > 0\} \qquad (3.16) \\
&\leq P\{f^*_{in}(y+1) - f^*_{in}(y)\frac{\beta(y+1)}{\beta(y)}[\varphi_i(y) + t] > 0 \text{ for some } y \leq x\} \\
&= P\{[F^*_{in}(y+1) - F^*_{in}(y)] - [F^*_{in}(y) - F^*_{in}(y-1)]\frac{\beta(y+1)}{\beta(y)}[\varphi_i(y) + t] > 0 \text{ for some } y \leq x\} \\
&\leq 3P\{\sup_{y \geq 0}|F^*_{in}(y) - F_i(y)| \geq \min_{y \leq x} \frac{|f_i(y+1) - f_i(y)\frac{\beta(y+1)}{\beta(y)}[\varphi_i(y) + t]|}{3[1 + \frac{\beta(y+1)}{\beta(y)}[\varphi_i(y) + t]}\},
\end{aligned}$$

which tends to zero as n tends to infinity.

Based on the above discussions, we have shown that $\varphi^*_{in}(x) \xrightarrow{P} \varphi_i(x)$ for each $x = 0, 1, \ldots$ and each $i = (0), 1, \ldots, k$. Therefore, the sequence of empirical Bayes selection rules $\{d^*_n\}$ is asymptotically optimal.

4. Asymptotic Optimality of the Monotone Estimators

In this section, we study the asymptotic optimality property of the estimators $\varphi^*_{in}(x)$. It is known that under the squared error loss, $\varphi_i(x)$ is the Bayes estimator of p_i given $X_i = x$. The associated Bayes risk is

$$R_i(G_i) = E[(P_i - \varphi_i(X_i))^2]. \qquad (4.1)$$

Let $\psi_i(\cdot)$ be any estimator of p_i with the associated Bayes risk $R_i(G_i, \psi_i)$. Then

$$R_i(G_i, \psi_i) - R_i(G_i) = E[(\psi_i(X_i) - \varphi_i(X_i))^2]. \qquad (4.2)$$

Let $\{\psi_{in}(x; X_{i1}, \ldots, X_{in}) \equiv \psi_{in}(x)\}$ be a sequence of empirical Bayes estimators based on $(x; X_{i1}, \ldots, X_{in})$.

Definition 4.1.

a) A sequence of empirical Bayes estimators $\{\psi_{in}\}_{n=1}^{\infty}$ is said to be asymptotically optimal relative to the prior distribution G if $R_i(G_i,\ \psi_{in}) \to R_i(G_i)$ as $n \to \infty$.

b) A sequence of empirical Bayes estimators $\{\psi_{in}\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order α_n relative to the prior distribution G_i if $R_i(G_i,\ \psi_{in}) - R_i(G_i) \leq O(\alpha_n)$ as $n \to \infty$, where $\{\alpha_n\}$ is a sequence of positive values such that $\lim\limits_{n\to\infty} \alpha_n = 0$.

From Section 3, $\varphi_{in}^*(x) \xrightarrow{P} \varphi_i(x)$ for all $x = 0, 1, 2, \ldots$. Thus, $\{\varphi_{in}^*\}_{n=1}^{\infty}$ is asymptotically optimal. However, the usefulness of empirical Bayes estimators in practical applications clearly depends on the convergence rates with which the risks for the successive estimation problems approach the optimal Bayes risk. Hence in the following, we study the convergence rates of the sequence of empirical Bayes estimators $\{\varphi_{in}^*\}$.

For $0 < \varepsilon < 1$, let

$$A_i(\varepsilon) = \{x | f_i(x) < \varepsilon\}, \ B_i(\varepsilon) = \{x | f_i(x) \geq \varepsilon\}. \tag{4.3}$$

<u>Assumption A</u>:

A1. There exist $t_i \epsilon (0, 1]$ and a positive constant c_i such that $P(A_i(\varepsilon)) \leq c_i \varepsilon^{t_i}$ for all $\varepsilon \in (0, 1)$.

A2. There exists a positive integer N_i such that $f_i(x)$ is decreasing in x for $x \geq N_i$.

<u>Remark 4.1.</u>: An example such that Assumption A holds true is given in Lin (1972).

We have the following theorem.

Theorem 4.1. Let $\{\varphi_{in}^*\}_{n=1}^{\infty}$ be the sequence of empirical Bayes estimators defined in (3.9). Then, under Assumption A,

$$R_i(G_i, \varphi_{in}^*) - R_i(G_i) \leq O(n^{-t_i/(2+t_i)}).$$

The proof of Theorem 4.1 can be obtained based on the following arguments.

For the empirical Bayes estimator φ_{in}^*, straight computation leads to

$$
\begin{aligned}
0 &\leq R_i(G_i, \varphi_{in}^*) - R_i(G_i) \\
&= E[(\varphi_{in}^*(X_i) - \varphi_i(X_i))^2] \\
&= \sum_{x=0}^{\infty} E[(\varphi_{in}^*(x) - \varphi_i(x))^2 | X_i = x] f_i(x) \\
&= \sum_{x \in A_i(\delta_n)} E[(\varphi_{in}^*(x) - \varphi_i(x))^2 | X_i = x] f_i(x) \\
&\quad + \sum_{x \in B_i(\delta_n)} E[(\varphi_{in}^*(x) - \varphi_i(x))^2 | X_i = x] f_i(x),
\end{aligned}
\tag{4.4}
$$

where $\delta_n = n^{-\alpha_i}$ and $\alpha_i = \frac{1}{2+t_i}$.

Lemma 4.1. Under Assumption A,

$$\sum_{x \in A_i(\delta_n)} E[(\varphi_{in}^*(x) - \varphi_i(x))^2 | X_i = x] f_i(x) \leq O(n^{-t_i/(2+t_i)}).$$

Proof: Note that $0 \leq \varphi_{in}^*(x),\ \varphi_i(x) \leq 1$. Thus,

$$
\begin{aligned}
&\sum_{x \in A_i(\delta_n)} E[(\varphi_{in}^*(x) - \varphi_i(x))^2 | X_i = x] f_i(x) \\
&\quad \leq \sum_{x \in A_i(\delta_n)} f_i(x) \\
&\quad = P(A_i(\delta_n)) \\
&\quad \leq c_i n^{-\alpha_i t_i} \quad \text{(by the definition of } A_i(\delta_n) \text{ and Assumption } A1) \\
&\quad = O(n^{-t_i/(2+t_i)}).
\end{aligned}
$$

Thus, it suffices to study the asymptotic behavior of $\sum_{x \in B_i(\delta_n)} E[(\varphi_{in}^*(x) - \varphi_i(x))^2 | X_i = x] f_i(x)$. Now, for each $x \in B_i(\delta_n)$,

$$
\begin{aligned}
&E[(\varphi_{in}^*(x) - \varphi_i(x))^2 | X_i = x] \\
&= E[(\varphi_{in}^*(x) - \varphi_i(x))^2 I_{in}(x) | X_i = x] \\
&\quad + E[(\varphi_{in}^*(x) - \varphi_i(x))^2 (1 - I_{in}(x)) | X_i = x],
\end{aligned}
\tag{4.5}
$$

where

$$I_{in}(x) = \begin{cases} 1 & \text{if } f_{in}(x) > 0, \\ 0 & \text{if } f_{in}(x) = 0, \end{cases} \tag{4.6}$$

and $f_{in}(x)$ is as defined in (3.2).

Lemma 4.2. For $x \in B_i(\delta_n)$,

$$E[(\varphi_{in}^*(x) - \varphi_i(x))^2(1 - I_{in}(x))|X_i = x] \le O(n^{-t_i/(2+t_i)}).$$

Proof:

$$\begin{aligned} & E[(\varphi_{in}^*(x) - \varphi_i(x))^2(1 - I_{in}(x))|X_i = x] \\ & \le P\{f_{in}(x) = 0\} \\ & = P\{f_{in}(x) - f_i(x) < -f_i(x)\} \\ & \le \exp\{-2nf_i^2(x)\} \quad \text{(by Theorem 1 of Hoeffding (1963))} \\ & \le \exp\{-2n\delta_n^2\} \quad (\text{since } x \in B_i(\delta_n)) \\ & = \exp\{-n^{t_i/(2+t_i)}\} \\ & \le O(n^{-t_i/(2+t_i)}). \end{aligned}$$

<u>Remark 4.2.</u>: Note that in Lemma 4.2, the upper bound is independent of x for all $x \in B_i(\delta_n)$.

Next, a straight computation leads to:

$$\begin{aligned} & E[(\varphi_{in}^*(x) - \varphi_i(x))^2 I_{in}(x)|X_i = x] \\ & = \int_0^{\varphi_i(x)} 2sP\{\varphi_{in}^*(x) - \varphi_i(x) < -s, f_{in}(x) > 0\}ds \\ & \quad + \int_0^{1-\varphi_i(x)} 2sP\{\varphi_{in}^*(x) - \varphi_i(x) > s, f_{in}(x) > 0\}ds. \end{aligned} \tag{4.7}$$

Lemma 4.3. a) For $x \in B_i(\delta_n)$, and $s \in (0, \varphi_i(x))$,

$$P\{\varphi_{in}^*(x) - \varphi_i(x) < -s \text{ and } f_{in}(x) > 0\} \leq 3d_i \exp\{-2n[\frac{q(x,s)}{3(1+p(x,s))}]^2\}$$

for some positive constant d_i, where $p(x,s) = \frac{f_i(x+1)}{f_i(x)} - \frac{s\beta(x+1)}{\beta(x)} > 0$ and $q(x,s) = \frac{s\beta(x+1)f_i(x)}{\beta(x)} > 0$.

b) For $x \in B_i(\delta_n)$,

$$\int_0^{\varphi_i(x)} 2sP\{\varphi_{in}^*(x) - \varphi_i(x) < -s \text{ and } f_{in}(x) > 0\}ds \leq O(n^{-t_i/(2+t_i)})$$

and the upper bound is independent of x for all $x \in B_i(\delta_n)$.

Proof: a) First, it is trivial that $p(x,s) > 0$ and $q(x,s) > 0$. Next, that $f_{in}(x) > 0$ and from the definitions of $f_{in}^*(x)$ and $h_{in}^*(x)$, it follows that $f_{in}^*(x) > 0$ and $h_{in}^*(x) > 0$. Also, $\varphi_{in}^*(x) \geq \varphi_{in}(x)$, by the definition of $\varphi_{in}^*(x)$. Then, following (3.14), we obtain: For $x \in B_i(\delta_n)$ and $s \in (0, \varphi_i(x))$,

$$\begin{aligned}
&P\{\varphi_{in}^*(x) - \varphi_i(x) < -s \text{ and } f_{in}(x) > 0\} \\
&\leq P\{\varphi_{in}(x) - \varphi_i(x) < -s \text{ and } f_{in}(x) > 0\} \\
&\leq P\{h_{in}^*(x+1) - h_{in}^*(x)[\varphi_i(x) - s] < 0\} \\
&\leq 3P\{\sup_{y \geq 0} |F_{in}(y) - F_i(y)| \geq \frac{q(x,s)}{3[1+p(x,s)]}\} \\
&\leq 3d_i \exp\{-2n[\frac{q(x,s)}{3[1+p(x,s)]}]^2\},
\end{aligned}$$

for some positive constant d_i, where the last inequality follows from Lemma 2.1 of Schuster (1968).

b) By using the fact that $0 < \frac{h_i(x+1)}{h_i(x)} < 1$ and $\frac{\beta(x+1)}{\beta(x)} \leq r$ for all $x \geq 0$, we have $1 + p(x,s) \leq 1 + r$. Then, from the result of part a) of this Lemma, we obtain

$$\begin{aligned}
&\int_0^{\varphi_i(x)} 2sP\{\varphi_{in}^*(x) - \varphi_i(x) < -s, f_{in}(x) > 0\}ds \\
&\leq \int_0^{\varphi_i(x)} 6d_i\ s\ \exp\{-\frac{2n}{9}[\frac{sf_i(x)\frac{\beta(x+1)}{\beta(x)}}{1+r}]^2\}ds \\
&\leq 13.5d_i\frac{(1+r)^2\beta^2(x)}{\beta^2(x+1)} \times \frac{1}{nf_i^2(x)} \\
&\leq 13.5d_i(1+r)^2\frac{1}{n\delta_n^2} \quad \text{(since } x \in B_i(\delta_n)) \text{ and therefore, } f_i(x) \geq \delta_n) \\
&= O(n^{-t_i/(2+t_i)}).
\end{aligned}$$

Note this upper bound is independent of $x \in B_i(\delta_n)$.

By Assumption A2, $f_i(x)$ is decreasing in x for all $x \geq N_i$. In the following, we only consider the case where n is large enough such that $\delta_n \equiv n^{-\frac{1}{2+t_i}} \leq f_i(y)$ for all $y \leq N_i$. Thus as $x \in B_i(\delta_n)$, then $f_i(y) \geq \delta_n$ for all $y \leq x$ (this holds true for either $x \leq N_i$ or $x > N_i$). Therefore, analogous to (3.16), by the definition of $\varphi_{in}^*(x)$, we obtain:

For $s \in (0, 1 - \varphi_i(x))$,

$$\begin{aligned}
&P\{\varphi_{in}^*(x) - \varphi_i(x) > s, f_{in}(x) > 0\} \\
&= P\{\varphi_{in}(y) - \varphi_i(x) > s \text{ for some } y \leq x, f_{in}(x) > 0\} \\
&\leq 3P\{\sup_{y \geq 0}|F_{in}^*(y) - F_i(y)| \geq -\min_{y \leq x} H(y)\} \\
&\leq 3d_i\ \exp\{-2n[-\min_{y \leq x} H(y)]^2\} \text{ (by Lemma 2.1 of Schuster (1968))}, \qquad (4.8)
\end{aligned}$$

where

$$H(y) = \frac{f_i(y+1) - f_i(y)\frac{\beta(y+1)}{\beta(y)}[\varphi_i(y) + s]}{3[1 + \frac{\beta(y+1)}{\beta(y)}(\varphi_i(y) + s)]}$$

$$= \frac{-sf_i(y)\frac{\beta(y+1)}{\beta(y)}}{3[1+\frac{\beta(y+1)}{\beta(y)}]}$$

$$\leq \frac{-sf_i(y)}{3[1+r]} \tag{4.9}$$

$$\leq \frac{-s\delta_n}{3[1+r]} \quad (\text{ since } f_i(y) \geq \delta_n)$$

$$< 0.$$

Lemma 4.4. For n sufficiently large, and $x \in B_i(\delta_n)$,

$$\int_0^{1-\varphi_i(x)} 2sP\{\varphi_{in}^*(x) - \varphi_i(x) > s, f_{in}(x) > 0\}ds \leq O(n^{-t_i/(2+t_i)}).$$

Proof: From (4.8) and (4.9), for n sufficiently large, as $x \in B_i(\delta_n)$,

$$\int_0^{1-\varphi_i(x)} 2sP\{\varphi_{in}^*(x) - \varphi_i(x) > s, f_{in}(x) > 0\}ds$$

$$\leq \int_0^{1-\varphi_i(x)} 6sd_i \exp\{-\frac{2n\delta_n^2 s}{9(1+r)^2}\}ds$$

$$= O(n^{-t_i/(2+t_i)}).$$

From Lemmas 4.2, 4.3, 4.4 and (4.7), we have: For $x \in B_i(\delta_n)$,

$$E[(\varphi_{in}^*(x) - \varphi_i(x))^2 | X_i = x] \leq O(n^{-t_i/(2+t_i)}).$$

This upper bound is independent of $x \in B_i(\delta_n)$. Therefore, we conclude that

$$\sum_{x \in B_i(\delta_n)} E[(\varphi_{in}^*(x) - \varphi_i(x))^2 | X_i = x] f_i(x) \leq O(n^{-t_i/(2+t_i)}). \tag{4.10}$$

Then, Lemma 4.1, (4.4) and (4.10) together complete the proof of Theorem 4.1.

5. Asymptotic Optimality of the Empirical Bayes Selection Rules

In this section, we study the asymptotic optimality of the sequence of empirical Bayes selection rules $\{d_n^*\}$.

Let $\{d_n\}_{n=1}^{\infty}$ be a sequence of empirical Bayes selection rules relative to the prior distribution G. Let $r(G, d_n)$ denote the expected Bayes risk of the selection rule d_n. Since the Bayes rule d_G achieves the minimum Bayes risk $r(G)$, $r(G, d_n) - r(G) \geq 0$ for all $n = 1, 2, \ldots$. Thus, the nonnegative difference $r(G, d_n) - r(G)$ is used as a measure of the optimality of the sequence of empirical Bayes selection rules $\{d_n\}$.

Definition 5.1. The sequence of empirical Bayes selection rules $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order α_n relative to the prior distribution G is $r(G, d_n) - r(G) \leq O(\alpha_n)$ as $n \to \infty$, where $\{\alpha_n\}$ is a sequence of positive numbers such that $\lim\limits_{n \to \infty} \alpha_n = 0$.

In the following, we evaluate the asymptotic behavior of the sequence of empirical Bayes rules $\{d_n^*\}$ proposed in Section 3, according to whether the control parameter p_0 is known or unknown.

<u>Convergence Rates of $\{d_n^*\}$ for p_0 Known Case</u>

For each $i = 1, \ldots, k$, let $S_i = \{x | \varphi_i(x) < p_0\}$ and $T_i = \{x | \varphi_i(x) > p_0\}$. Define

$$M_i = \begin{cases} \min T_i & \text{if } T_i \neq \phi, \\ \infty & \text{if } T_i = \phi, \end{cases} \tag{5.1}$$

and

$$m_i = \begin{cases} \max S_i & \text{if } S_i \neq \phi, \\ -1 & \text{if } s_i = \phi. \end{cases} \tag{5.2}$$

By the increasing property of $\varphi_i(x)$ with respect to the variable $x = 0, 1, \ldots, m_i \leq M_i$; also $m_i < M_i$ if $T_i \neq \phi$. Furthermore,

$$\begin{cases} x \leq m_i & \text{iff } \varphi_i(x) < p_0; \text{ and} \\ y \geq M_i & \text{iff } \varphi_i(y) > p_0. \end{cases} \tag{5.3}$$

Note that when $T_i = \phi$, it meas that in terms of its quality, the population π_i is bad. Also, for $S_i = \phi$, it means that in terms of its quality population π_i is as good as the control. We exclude these extreme cases, in the following, and study the asymptotic behavior of the sequence of empirical Bayes selection rules $\{d_n^*\}$ under Assumption B.

Assumption B: $T_i \neq \phi$ and $S_i \neq \phi$ for all $i = 1, \ldots, k$.

Now, for the empirical Bayes selection rule d_n^*, a direct computation leads to the following:

$$\begin{aligned} 0 &\leq r(G, d_n^*) - r(G) \\ &= \sum_{i=1}^{k} D_i(G, d_n^*), \end{aligned} \tag{5.4}$$

where

$$D_i(G, d_n^*) = \sum_{x=0}^{m_i} [p_0 - \varphi_i(x)] P\{\varphi_{in}^*(x) \geq p_0\} f_i(x) + \sum_{x=M_i}^{\infty} [\varphi_i(x) - p_0] P\{\varphi_{in}^*(x) < p_0\} f_i(x). \tag{5.5}$$

By the nondecreasing property of the estimator φ_{in}^*, we have

$$\begin{cases} P\{\varphi_{in}^*(x) \geq p_0\} \leq P\{\varphi_{in}^*(m_i) \geq p_0\} & \text{for all } x \leq m_i, \text{ and} \\ P\{\varphi_{in}^*(y) < p_0\} \leq P\{\varphi_{in}^*(M_i) < p_0\} & \text{for all } y \geq M_i. \end{cases} \tag{5.6}$$

Thus, from (5.5) and (5.6),

$$D_i(G, d_n^*) \leq P\{\varphi_{in}^*(m_i) \geq p_0\} b_{i1} + P\{\varphi_{in}^*(M) < p_0\} b_{i2}, \tag{5.7}$$

where $b_{i1} = \sum_{x=0}^{m_i} [p_0 - \varphi_i(x)] f_i(x)$, and $b_{i2} = \sum_{x=M_i}^{\infty} [\varphi_i(x) - p_0] f_i(x)$. Note that $0 \leq b_{i1},\ b_{i2} \leq 1$.

Now,

$$\begin{aligned}
&P\{\varphi_{in}^*(M_i) < p_0\} \\
&= P\{\varphi_{in}^*(M_i) < p_0,\ \max(X_{i1}, \ldots, X_{in}) \leq M_i\} \\
&\quad + P\{\varphi_{in}^*(M_i) < p_0,\ \max(X_{i1}, \ldots, X_{in}) > M_i\} \qquad (5.8) \\
&\leq [F_i(M_i)]^n + P\{\varphi_{in}^*(M_i) < p_0,\ \max(X_{i1}, \ldots, X_{in}) > M_i\}.
\end{aligned}$$

Analogous to Lemma 4.3. a), we can obtain

$$\begin{aligned}
&P\{\varphi_{in}^*(M_i) < p_0 \text{ and } \max\ (X_{i1}, \ldots, X_{in}) > M_i\} \\
&\leq P\{h_{in}^*(M_i + 1) < h_{in}^*(M_i) p_0 \text{ and } \max(X_{i1}, \ldots, X_{in}) > M_i\} \\
&\leq 3P\{\sup_{y \geq 0} |F_{in}^*(y) - F_i(y)| > \frac{\Delta(M_i, p_0)}{3[1 + \frac{\beta(M_i+1)}{\beta(M_i)} p_0]}\} \\
&\leq 3d_i \exp\{-\frac{2}{9} n \Delta^2(M_i, p_0)/[1 + \frac{\beta(M_i + 1)}{\beta(M_i)} p_0]^2\}, \qquad (5.9)
\end{aligned}$$

where $\Delta(M_i, p_0) = f_i(M_i + 1) - f_i(M_i) \frac{\beta(M_i+1)}{\beta(M_i)} p_0 > 0$.

From (5.8) and (5.9), we have

$$P\{\varphi_{in}^*(M_i) < p_0\} \leq O(\exp(-\tau_{i1} n)) \qquad (5.10)$$

where $\tau_{i1} = \min\left(\frac{2\Delta^2(M_i, p_0)}{9[1 + \frac{\beta(M_i+1)}{\beta(M_i)} p_0]^2},\ \ell n \frac{1}{F_i(M_i)}\right)$.

Also,

$$\begin{aligned}
&P\{\varphi_{in}^*(m_i) \geq p_0\} \\
&= P\{\varphi_{in}^*(m_i) \geq p_0, \max(X_{i1}, \ldots, X_{in}) \leq m_i\} \\
&\quad + P\{\varphi_{in}^*(m_i) \geq p_0,\ \max(X_{i1}, \ldots, X_{in}) > m_i\} \\
&\leq [F_i(m_i)]^n + P\{\varphi_{in}^*(m_i) \geq p_0, \max(X_{i1} \ldots X_{in}) > m_i\}. \qquad (5.11)
\end{aligned}$$

Analogous to (3.16), we obtain

$$
\begin{aligned}
&P\{\varphi_{in}^*(m_i) \geq p_0,\ \max(X_{i1}\dots X_{in}) > m_i\} \\
&= P\{\varphi_{in}(y) \geq p_0 \text{ for some } y \leq m_i, \max(X_{i1}\dots X_{in}) > m_i\} \\
&\leq 3P\{\sup_{x\geq 0}|F_{in}^* - F_i(x)| \geq \min_{y\leq m_i} \frac{|\Delta(y,p_0)|}{3[1+\frac{\beta(y+1)}{\beta(y)}p_0]}\} \\
&\leq 3d_i \exp\left\{-\frac{2n}{9}\min_{y\leq m_i}\left[\frac{|\Delta(y,p_0)|}{1+\frac{\beta(y+1)}{\beta(y)}p_0}\right]^2\right\},
\end{aligned} \tag{5.12}
$$

where $\Delta(y,p_0) = f_i(y+1) - f_i(y)\frac{\beta(y+1)}{\beta(y)}p_0$. Note that $\Delta(y,p_0) < 0$ for all $y \leq m_i$.

Thus,

$$
P\{\varphi_{in}^*(m_i) \geq p_0\} \leq O(\exp\{-\tau_{i2}n\}), \tag{5.13}
$$

where $\tau_{i2} = \min\left(\frac{2}{9}\min_{y\leq m_i}\left\{\frac{|\Delta(y,p_0)|^2}{[1+\frac{\beta(y_1)}{\beta(y)}p_0]^2}\right\}, \ell n\frac{1}{F_i(m_i)}\right)$.

Let $\tau_i = \min(\tau_{i1}, \tau_{i1})$, and $\tau = \min(\tau_1,\dots,\tau_k)$. Note that $\tau > 0$, since $\tau_i > 0$ for each $i = 1,\dots,k$. From the above results the following theorem follows:

Theorem 5.1. Under Assumption B, we have:

a) $D_i(G, d_n^*) \leq O(\exp(-\tau_i n))$ for each $i = 1\dots k$, and

b) $r(G, d_n^*) - r(G) \leq O(\exp(-\tau n))$.

Convergence Rates of $\{d_n^*\}$ for p_0 Unknown Case

When the parameter p_0 is unknown, the convergence rates of the sequence of empirical Bayes selection rules $\{d_n^*\}_{n=1}^{\infty}$ is evaluated under Assumption A which is given in Section 4. Without loss of generality, in this section, we assume that $c_0 = c_1 = \dots = c_k = c > 0$ and $t_0 = t_1 = \dots = t_k = t \in (0,1]$, where the parameters c_i, t_i, $i = 0,1\dots k$ are given in Assumption A1.

For each $i = 1, \ldots, k$, let

$$S_i = \{(x_0, x_i) | \varphi_i(x_i) < \varphi_0(x_0)\}, T_i = \{(x_0, x_i) | \varphi_i(x_i) \geq \varphi_0(x_0)\}$$

$$E_{in} = \{(x_0, x_i) || \varphi_i(x_i) - \varphi_0(x_0)| \leq \varepsilon_n\}, \text{ and}$$

$$E_{in}^c = \{(x_0, x_i) || \varphi_i(x_i) - \varphi_0(x_0)| > \varepsilon_n\} \text{ where } \varepsilon_n > 0.$$

Thus,

$$\begin{aligned} 0 &\leq r(G, d_n^*) - r(G) \\ &= \sum_{i=1}^{K} D_i^*(G, d_n^*), \end{aligned} \tag{5.14}$$

where

$$\begin{aligned} D_i^*(G, d_n^*) = &\sum_{(x_0,x_i) \in T_i} Q_i(x_0, x_i) + \sum_{(x_0,x_i) \in S_i} R_i(x_0, x_i) \\ = &\sum_{T_i \cap E_{in}} Q_i(x_0, x_i) + \sum_{T_i \cap E_{in}^c \cap A_0(\delta_n) \cap A_i(\delta_n)} Q_i(x_0, x_i) + \sum_{T_i \cap E_{in}^c \cap A_0(\delta_n) \cap B_i(\delta_n)} Q_i(x_0, x_i) \\ &+ \sum_{T_i \cap E_{in}^c \cap B_0(\delta_n) \cap A_i(\delta_n)} Q_i(x_0, x_i) + \sum_{T_i \cap E_{in}^c \cap B_0(\delta_n) \cap B_i(\delta_n)} Q_i(x_0, x_i) \\ &+ \sum_{S_i \cap E_{in}} R_i(x_0, x_i) + \sum_{S_i \cap E_{in}^c \cap A_0(\delta_n) \cap A_i(\delta_n)} R_i(x_0, x_i) + \sum_{S_i \cap E_{in}^c \cap A_0(\delta_n) \cap B_i(\delta_n)} R_i(x_0, x_i) \\ &+ \sum_{S_i \cap E_{in}^c \cap B_0(\delta_n) \cap A_i(\delta_n)} R_i(x_0, x_i) + \sum_{S_i \cap E_{in}^c \cap B_0(\delta_n) \cap B_i(\delta_n)} R_i(x_0, x_i), \end{aligned} \tag{5.15}$$

where

$$\begin{cases} Q_i(x_0, x_i) = [\varphi_i(x_i) - \varphi_0(x_0)] P\{\varphi_{in}^*(x_i) < \varphi_{0n}^*(x_0)\} f_i(x_i) f_0(x_0), \\ R_i(x_0, x_i) = [\varphi_0(x_0) - \varphi_i(x_i)] P\{\varphi_{in}^*(x_i) \geq \varphi_{0n}^*(x_0)\} f_i(x_i) f_0(x_0), \end{cases} \tag{5.16}$$

and the notation $\sum\limits_{A}$ means that the summation is computed over the set A.

Careful examination leads to the following results:

$$\sum_{T_i \cap E_{in}} Q_i(x_0, x_i) \leq O(\varepsilon_n), \sum_{S_i \cap E_{in}} R_i(x_0, x_i) \leq O(\varepsilon_n),$$

$$\sum_{T_i \cap E_{in}^c \cap A_0(\delta_n) \cap A_i(\delta_n)} Q_i(x_0, x_i) \le O(\delta_n^{2t}),$$

$$\sum_{S_i \cap E_{in}^c \cap A_0(\delta_n) \cap A_i(\delta_n)} R_i(x_0, x_i) \le O(\delta_n^{2t}), \tag{5.17}$$

$$\sum_{T_i \cap E_{in}^c \cap A_0(\delta_n) \cap B_i(\delta_n)} Q_i(x_0, x_i) \le O(\delta_n^{t}),$$

$$\sum_{S_i \cap E_{in}^c \cap A_0(\delta_n) \cap B_i(\delta_n)} R_i(x_0, x_i) \le O(\delta_n^{t}),$$

$$\sum_{T_i \cap E_{in}^c \cap B_0(\delta_n) \cap A_i(\delta_n)} Q_i(x_0, x_i) \le O(\delta_n^{t}), \text{ and}$$

$$\sum_{S_i \cap E_{in}^c \cap B_0(\delta_n) \cap A_i(\delta_n)} R_i(x_0, x_i) \le O(\delta_n^{t}).$$

Now, for $(x_0, x_i) \in T_i \cap E_{in}^c \cap B_0(\delta_n) \cap B_i(\delta_n)$, $f_i(x_i) \ge \delta_n$, $f_0(x_0) \ge \delta_n$, and $\varepsilon_n < \varphi_i(x_i) - \varphi_0(x_0) < 1$. Thus,

$$\begin{aligned} & P\{\varphi_{in}^*(x_i) < \varphi_{0n}^*(x_0)\} \\ &= P\{[\varphi_{in}^*(x_i) - \varphi_i(x_i)] - [\varphi_{0n}^*(x_0) - \varphi_0(x_0)] < \varphi_0(x_0) - \varphi_i(x_i)\} \\ &\le P\{\varphi_{in}^*(x_i) - \varphi_i(x_i) < -\frac{\varepsilon_n}{2}\} + P\{\varphi_{0n}^*(x_0) - \varphi_0(x_0) > \frac{\varepsilon_n}{2}\}. \end{aligned} \tag{5.18}$$

Now,

$$\begin{aligned} & P\{\varphi_{in}^*(x_i) - \varphi_i(x_i) < -\frac{\varepsilon_n}{2}\} \\ &= P\{\varphi_{in}^*(x_i) - \varphi_i(x_i) < -\frac{\varepsilon_n}{2}, f_{in}(x_i) = 0\} + P\{\varphi_{in}^*(x_i) - \varphi_i(x_i) < -\frac{\varepsilon_n}{2}, f_{in}(x_i) > 0\}, \end{aligned}$$

where

$$\begin{aligned} & P\{\varphi_{in}^*(x_i) - \varphi_i(x_i) < -\frac{\varepsilon_n}{2}, f_{in}(x_i) = 0\} \\ &\le P\{f_{in}(x_i) = 0\} \\ &\le \exp\{-2n\delta_n^2\} (\text{ since } x_i \in B_i(\delta_n); \text{ and see Lemma 4.2}), \end{aligned}$$

and

$$\begin{aligned} & P\{\varphi_{in}^*(x_i) - \varphi_i(x_i) < -\frac{\varepsilon_n}{2}, f_{in}(x_i) > 0\} \\ &\le 3d_i \exp\{-\frac{\beta^2(x_i+1)}{18(1+r)^2\beta^2(x_i)} n\varepsilon_n^2\delta_n^2\} \\ &\le O(\exp\{-\frac{n\varepsilon_n^2\delta_n^2}{18(1+r)^2}\}). \text{ (The proof is analogous to that of Lemma 4.3.a).} \end{aligned}$$

Therefore,

$$P\{\varphi_{in}^*(x_i) - \varphi_i(x_i) < -\frac{\varepsilon_n}{2}\} \leq O(\exp\{-\frac{n\varepsilon_n^2\delta_n^2}{18(1+r)^2}\}). \tag{5.19}$$

Next, under Assumption A2, along the line of (4.8) and the argument given there, for n sufficiently large, we have

$$P\{\varphi_{0n}^*(x_0) - \varphi_0(x_0) > \frac{\varepsilon_n}{2}\} \leq O(\exp\{-\frac{n\varepsilon_n^2\delta_n^2}{18(1+r)^2}\}). \tag{5.20}$$

Note that the convergence rates obtained at (5.19) and (5.20) are independent of $(x_0, x_i) \in T_i \cap E_{in}^c \cap B_0(\delta_n) \cap B_i(\delta_n)$. Therefore, from (5.16) and (5.18),

$$\sum_{T_i \cap E_{in}^c \cap B_0(\delta_n) \cap B_i(\delta_n)} Q_i(x_0, x_i) \leq O(\exp\{-\frac{n\varepsilon_n^2\delta_n^2}{18(1+r)^2}\}). \tag{5.21}$$

Similarly, we can also conclude that

$$\sum_{S_i \cap E_{in}^c \cap B_0(\delta_n) \cap B_i(\delta_n)} R_i(x_0, x_i) \leq O(\exp\{-\frac{n\varepsilon_n^2\delta_n^2}{18(1+r)^2}\}). \tag{5.22}$$

By letting $\delta_n = \left[\frac{18(1+r)^2 \ell n\ n}{(2+2t)n}\right]^{\frac{1}{2+2t}}$ and $\varepsilon_n = \delta_n^t$, we have the following theorem.

Theorem 5.2. Under Assumption A, we have

a) $D_i^*(G, d_n^*) \leq O(\delta_n^t)$ for $i = 1, ..., k$, and

b) $r(G, d_n^*) - r(G) \leq O(\delta_n^t)$.

Proof: The proof follows directly from (5.14), (5.15), (5.17), (5.21) and (5.22).

References

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

Deely, J. J. (1965). Multiple decision procedures from an empirical Bayes approach. Ph.D. Thesis (Mimeo. Ser. No. 45), Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *J. Statist. Plann. Infer.* **8**, 87–101.

Gupta, S. S. and Leu, L. Y. (1983). On Bayes and empirical Bayes rules for selecting good populations. Technical Report #83–37, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. and Liang, T. (1986a). Empirical Bayes rules for selecting good binomial populations. *Adaptive Statistical Procedures and Related Topics* (Ed. J. Van Ryzin), IMS Lecture Notes-Monograph Series, **Vol. 8**, 110–128.

Gupta, S. S. and Liang, T. (1986b). Parametric empirical Bayes rules for selecting the most probable multinomial event. Submitted for publication.

Gupta, S. S. and Liang, T. (1987). Empirical Bayes rules for selecting the best binomial population. *Statistical Decision Theory and Related Topics-IV* (Eds. S. S. Gupta and J. O. Berger), Springer-Verlag, **Vol. I**, 213–224.

Gupta, S. S. and Liang, T. (1988). Selecting the best binomial population: parametric empirical Bayes approach. Submitted for publication.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **8**, 87–101.

Liang, T. (1988). On the convergence rates of empirical Bayes rules for two-action problems: discrete case. To appear in *Ann. Statist.*

Lin, P-E. (1972). Rates of convergence in empirical Bayes estimation problems: discrete case. *Ann. Inst. Statist. Math.* **23**, 319–325.

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.*, **1**, University of California Press, 157–163.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1–20.

Robbins, H. (1983). Some thoughts on empirical Bayes estimation. *Ann. Statist.* **11**, 713–723.

Schuster, E. F. (1968). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* **40**, 1187–1195.